Scientific Journal of Impact Factor(SJIF): 3.134

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 11, November -2015

Review paper on protein structure prediction using Various Selection Methods in Genetic Algorithm

CHANDRA AKASH KIRAN [1], MR. DEEPAK XAXA [2]

[1]School of Engineering and IT, MATS University, Raipur, India
[2]School of Engineering and IT, MATS University, Raipur, India

Abstract — *predicting the native structure of a protein from its amino acid sequence is one of the most challenging problems in bioinformatics. In protein structure prediction there are two important issues. The first one is the design of the structure model and the second one is the design of the optimization technology. As protein structure prediction problem has been proved to be an NP Hard problem. Thus meta-heuristic techniques used to solve the global optimization problem. In our study, we are going to compare performance of genetic algorithm with different selection method and crossover method for protein structure prediction.*

Keywords- Genetic Algorithm, Selection, protein structure prediction.

I. Introduction

Protein structure prediction is defined as the prediction of the tertiary structure of a protein by using its primary structure information. It has become an important research topic in bioinformatics and it has important applications in medicine and other fields, such as drug design, prediction of diseases, and so on. Because of the complexity of the realistic protein structure, it is hard to determine the exact tri-dimensional structure from its sequence of amino acids [1]. Therefore, a lot of coarse structure models have been developed. The HP model is the most conventional one among them and has been widely used in protein structure prediction. Different from the complex structure models, HP model only assumes two types of amino acids-hydrophobic (H) and hydrophilic (P) and the sequence of amino acids is assumed to be embedded in a lattice, which is used to discrete the space of conformations. For simplicity, the only interaction considered in HP model is the interaction between the nonadjacent but next-neighbored hydrophobic monomers, which is used to force the formation of a compact hydrophobic core as observed in real proteins [2]. Although simplified models have the capability of catching nontrivial aspects of the folding problem, the approximations involved are not really suitable [3].

The main reason lies in that local interactions are neglected in the simplified models. As is well known, local interactions might be important for the local structure of the chains [4] and no sequences with compact, well-defined native structures could be found if local interactions are neglected [3]. Therefore, many other models which consider local interactions have drawn a lot of attention and been proposed. The AB off-lattice model is the one that could meet the aforementioned requirement. Currently, AB off-lattice model has been widely applied to protein structure prediction and many improved models have been proposed based on the original model. In AB off-lattice model, two types of monomers are taken into consideration. The hydrophobic monomers are labeled by A while the hydrophilic ones are labeled by B. Different from HP model, the interactions considered in AB model include both sequence independent local interactions and the sequence dependent Lennard - Jones term that favors the formation of a hydrophobic core. After a structure model is adopted, an important issue in PSP is to develop an optimization technology to find the best conformation of a protein sequence based on the assumed structure model. However, protein structure prediction (PSP) is an NP-hard problem even when the simplest models are assumed [5, 6].
In order to tackle this issue, many heuristic approaches have been developed. In the past decades, researchers have developed many algorithms to solve the global optimization problem in protein folding structure prediction (PFSP).

II. METHODOLOGY

2.1 Population

Cleary [12] described the population as a result of a single iteration of genetic algorithm. Iteration can create a new population. Population contains a set of chromosomes; each chromosome is one complete possible solution to the problem to be solved using genetic algorithm.

2.2 Evaluations

For each chromosome there is a fitness function used to evaluate the fitness of each chromosome. Fitness's value reflects the quality of each chromosome.

2.3 Encoding

The gene is a problem parameter; it can be encoded as a binary, integer, or float number.

2.4 Selections

It is the process of selecting the chromosomes to apply Steady State Genetic Algorithm. Types of selection are:

2.4.1 Rank Selection

Rank Selection ranks the population and every chromosome receives fitness from the ranking. The worst has fitness 1 and the best has fitness N. It results in slow convergence but prevents too quick convergence. It also keeps up selection pressure when the fitness variance is low [13]. It preserves diversity and hence leads to a successful search. In Linear Rank selection, individuals are assigned subjective fitness based on the rank within the population. The individuals in the population are sorted from best to worst according to their fitness values. Each individual in the population is assigned a numerical rank based on fitness, and selection is based on this ranking rather than differences in fitness.

2.4.2 Elitism Selection

The idea here is to arrange the chromosomes in the decreasing order according to their fitness values. Then apply the selection with each two chromosomes in the arranged set. In this way, Genetic Algorithm will be applied between strong chromosomes or between weak chromosomes. This means there is no chance to apply Genetic Algorithm between weak and strong chromosomes [14]. Elitism is a kind of selection in which the best individual passed to the next generation as such without any modification. Elitism prevents the best individual to undergo the reproduction process so as to pass them without any modification into next generation.

2.4.3 Tournament Selection

GAs uses a selection mechanism to select individuals from the population to insert into a mating pool. Individuals from the mating pool are used to generate new offspring, with the resulting offspring forming the basis of the next generation. A selection mechanism in GA is simply a process that favors the selection of better individuals in the population for the mating pool. The selection pressure is the degree to which the better individuals are favored: the higher the select ion pressure, the more the better individuals are favored. This selection pressure drives the GA to improve the population fitness over succeeding generations. The convergence rate of a GA is largely determined by the selection pressure, with higher selection pressures resulting in higher convergence rates. However, if the selection pressure is too low, the convergence rate will be slow, and the GA will unnecessarily take longer to find the optimal solution. If the selection pressure is too high, there is an increased chance of the GA prematurely converging to an incorrect (suboptimal) solution. Tournament selection provides selection pressure by holding a tournament among s competitors, with s being the tournament size. The winner of the tournament is the individual with the highest fitness of the s tournament competitors. The winner is then inserted into the mating pool. The mating pool, being comprised of tournament winners, has a higher average fitness than the average population fitness. This fitness difference provides the selection pressure, which drives the GA to improve the fitness of each succeeding generation. Increased selection pressure can be provided by simply increasing the tournament size s, as the winner from a larger tournament will, on average, have a higher fitness than the winner of a smaller tournament [15].

2.5 Crossover

This process is used to interchange genes between chromosomes to create offspring. Types of crossover are:

2.5.1 Single Point

Select the crossover point within a chromosome randomly and interchange the two parent chromosomes at this point to produce two new offspring's.

2.5.2 Two Points

Select two points randomly and interchange the two parent genes between these points.

2.5.3 Uniform

According to some probability, crossover will decide the parent contribution in the offspring chromosome. If the mixing ratio is equal to 0.5 this means 50% of genes in the offspring will come from parent 1 and the other will come from parent 2.

2.6 Mutation

This process will change the value of randomly selected gene. Types of mutation are:

2.6.1 Flip Bit (Used for binary represented genes)

Choose one gene randomly and Flip the value of the chosen gene.

2.6.2 Boundary (Used for integer and float represented genes)

Choose one gene randomly and Replace the value of the gene with the upper or the lower value.

2.6.3 Uniform (Used for integer and float representation)

Choose one gene randomly and Replace the value of a chosen gene with a uniform random value selected between the user specified upper and lower bounds for that gene.

2.7 Replacements

This process will compare between several chromosomes to choose the best. Types of replacement are:

2.7.1 Binary Tournament:

It will take two chromosomes and according to their fitness function it will choose the best of them, and ignore the second one.

2.7.2 Triple Tournament:

It will replace the worst two chromosomes between three chromosomes by the chromosome with the highest fitness value.

2.8 Stopping Criterions

Starting with an initial population, the evolution process is repeated until the satisfaction of the end condition. Kumar et al. [17] mentioned common terminating conditions such as:

- The found solution satisfies the minimum criterion.
- A fixed number of generations reached.
- Allocating budget (ex: time, money) reached.
- Successive iterations no longer produce better results.

III CONCLUSION

This paper presented the three types of selection operators and three type of crossover operators in the Genetic algorithm. To find out the performance of selection operators and its time complexity for the best result.

REFERENCES

[1] Lopes HS "Evolutionary Algorithms for the Protein Folding Problem: A Review and Current Trends. Studies in Computational Intelligence" Springer Berlin 2008, 151:297-315

[2] Hart WE, Newman A "Protein structure prediction with lattice models" Handbook of Molecular Biology CRC Press Aluru S. Chapman & Hall/CRC Computer and Information Science Series 2006, 1-24

[3] Irbäck A, Sandelin E "Local Interactions and Protein Folding: Model Study on the Square and Triangular Lattices" J. Chem. Phys. 1998, 108(5):2245-2250

[4] Irbäck A, Peterson C, Potthast F, Sommelius O "Local interactions and protein folding: A three-dimensional off-lattice approach" J. Chem. Phys. 1997, 107:273-282

[5] Hart WE, Istrail S "Robust proofs of NP-hardness for protein folding general lattices and energy potentials" Journal of Computational Biology 1997, 4(1):1-22

[6] Ngo JT, Marks J, Karplus M "Computational complexity, protein structure prediction, and the Levinthal paradox. The Protein folding problem and tertiary structure prediction" Mertz M, Grand ML. S Birkhauser 1994, 433-506

[7] Hoque MT, Chetty M, Dooley LS "A New Guided Genetic Algorithm for 2D Hydrophobic-Hydrophilic Model to Predict Protein Folding" IEEE Congress on Evolutionary Computation 2005

[8] Corne DW, Fogel GB "An Introduction to Bioinformatics for Computer Scientists. Evolutionary Computation in Bioinformatics" Elsevier India Fogel GB, Corne DW 2004, 3-18

[9] Takahashi O, Kita H, Kobayashi S "Protein Folding by a Hierarchical Genetic Algorithm" 4th Int. Symp. AROB. 1999, 19-22

[10] König R, Dandekar T: "Refined Genetic Algorithm Simulation to Model Proteins" Journal of Molecular Modeling

Springer Berlin 1999, 5:317-324

[11] Zhang X, Lin X, Wan C, Li T: “Genetic-Annealing Algorithm for 3D Off lattice Protein Folding Model” PAKDD workshops 2007, 4819:186-193

[12] Cleary, B.(2011). “Problems with crossover bias for binary string representations in genetic algorithms”. Master thesis. Californai state university. Long beach. Californa. United States.

[13] S.N.Sivanandam, S.N.Deepa, ”An Introduction to Genetic Algorithms”.

[14] Firas Alabsi,Reyadh Naoum, ”Comparison of Selection Methods and Crossover Operations using Steady State Genetic Based Intrusion Detection System”,an Journal of Emerging Trends in Computing and Information Sciences, VOL. 3, NO.7, July 2012.

[15] Brad L. Mille r, David E. Goldberg,” Genetic Algorithms, Tournament Selection, and the Effects of Noise”.

[16] Dr. Rajib Kumar Bhattacharjya,”Introduction To Genetic Algorithms”.

[17] Kumar, M., Husian, M., Upreti, N., Gupta, D. (2010). “Genetic algorithm: review and application” International Journal of Information Technology and Knowledge Management. Vol (2). No (2). Page 451.